FROM RESEARCH TO PRACTICE

Evaluating Randomized Trials of Screening

H. Gilbert Welch, MD, MPH, William C. Black, MD

A smoker requests a screening chest x-ray because he has read that it will lower his chances of developing lung cancer. Later, a female patient asks your opinion about screening mammography. And at the end of day, a patient returns his stool hemeoccult cards unused because he found obtaining the specimen distasteful. How should you respond to the foregoing scenarios? What data are available on these and other screening strategies?

Screening, the systematic examination of asymptomatic persons to detect and treat subclinical disease, has become the primary weapon in the battle to prevent disease. Currently the U.S. Preventive Services Task Force offers guidance on more than 50 screening tests,[1] and the ongoing development of genetic testing provides a hint of the boundless opportunities for screening in the future.[2] The potential for exponential growth alone provides a strong motivation for physicians to be able to critically interpret evaluations of screening.[3]

But there are other reasons to carefully scrutinize screening strategies. First are the unique implications of screening for the population as a whole. Target disorders are relatively rare, and to find disease, many people must be screened. Unforeseen risks, however rare or minor, are compounded as they apply to all who are screened, while benefits accrue to only a few. Second, cursory evaluations of screening are subject to powerful biases that almost always favor the more intensive screening options.[4] Finally, there is the implied pledge of prevention given individuals who are well.[5,6] While symptomatic patients generally ask for our help, with screening we are telling asymptomatic people that they need our help. For these people there is an implied promise of future disease prevention—namely, that the screening strategy works. In this article, we offer clinicians a framework to evaluate how realistic such a promise might be using the research evidence from randomized trials.

THE FRAMEWORK

Building on the framework developed by the Evidence-Based Medicine Working Group for assessing articles about therapy or prevention,[7] we consider three broad questions (Table 1). To illustrate the framework, we examine three large randomized trials investigating screening strategies for cancers of the lung, breast, and colon (Table 2). Our intent is to provide specific examples of the issues relevant to screening trials, not a comprehensive review of the evidence on screening for these cancers.

The Mayo Lung Study investigated the usefulness of a chest x-ray every 4 months in reducing lung cancer mortality among male smokers.[8,9] Control subjects received the then current standard of care at the Mayo Clinic—about half had annual chest films. The Health Insurance Plan (HIP) Breast Project studied the efficacy of annual mammography (two-view) in reducing breast cancer mortality.[10] Control subjects had little exposure to the technology, as screening mammography was not a covered benefit in the insurance plan. Finally, the Minnesota Colon Cancer Control Study examined the value of annual fecal occult blood testing in reducing colorectal cancer mortality.[11] Less than 2% of colorectal cancers in the control group were found by fecal occult blood testing, suggesting that control group contamination was minimal.

IS THE STUDY APPLICABLE TO MY PRACTICE?

Before embarking on a detailed review of any trial, it is useful to consider its relevance to one's own practice, as we will briefly discuss.

From Research to Practice, *a Journal series, presents articles to heighten the clinician's awareness of research and methodology issues that have direct relevance to practice. If you wish to submit a manuscript for consideration for this series, please contact Cynthia D. Mulrow, MD, MSc, Associate Editor, at mulrow@uthscsa.edu, or contact the* Journal of General Internal Medicine *at (215) 823-4471 to receive the appropriate guidelines.*

Received from the Department of Veterans Affairs Medical Center, White River Junction, Vt. (HGW); the Center for the Evaluative Clinical Sciences, Dartmouth Medical School, Hanover, NH (HGW, WCB); and the Norris Cotton Cancer Center and the Department of Radiology, Dartmouth-Hitchcock Medical Center, Lebanon, NH (WCB).

Dr. Welch is supported by a Veterans Affairs Career Development Award in health services research and development. Dr. Black is a Radiological Society of North America Research and Education Fund Scholar.

Address correspondence and reprint requests to Dr. Welch: VA Outcomes Group (111B), Department of Veterans Affairs Medical Center, White River Junction, VT 05009.

Table 1. Questions for Critical Readers to Consider When Reviewing a Randomized Trial of Screening

Is the study applicable to my practice?
To whom does the study apply?
Is the screening test available to me? (with comparable performance/interpretation?)
Has the technology of the testing changed since the study?
Are the results of the study valid?
Were the groups similar at the start of the trial?
Were outcome assessments "blinded" to the randomization?
Were patients analyzed in the groups to which they were randomized?
What were the results?
Results about the test
Are more cases of early disease detected in the screened group?
Are fewer cases of advanced disease detected in the screened group?
Results about treatment following the test
Has advanced disease been prevented in the screened group? Is disease-specific mortality lower in the screened group? What is the magnitude of the absolute risk reduction?
Is it clear that early detection and treatment produce a net gain? What is the effect on all-cause mortality?

To Whom Does the Study Apply?

To answer this question, one must review both the inclusion and exclusion criteria. As illustrated in Table 2, the inclusion criteria for randomized trials of screening are generally broad and straightforward (e.g., patient age and perhaps specific risk factors for the target disease). Exclusions, on the other hand, are focused. Individuals with signs and symptoms of the target disease or who have had a history of the disease are typically excluded. The combination of broad inclusion criteria and focused exclusion criteria means that most clinicians should find that the results apply to some portion of their practice.

Is the Test Available to Me?

Even if one's patients are similar to the population in a study, it is not relevant if the test being investigated is not performed locally. Although this is not a concern for chest x-rays, mammography, and fecal occult blood testing, it may well be for screening tests in the future (such as those using genetic material). Furthermore, just because the test can be obtained locally does not guarantee that it will be performed or interpreted as it was in the trial. Although making judgments about the local quality of the test and the skill of local interpreters is difficult, these factors are critical in determining whether the results of the study are applicable to one's own practice.

Has the Technology of Testing Changed Since the Study?

If the technology of testing has changed since the trial's onset, then so may the results. Of the three trials considered here, this concern is most relevant for mammography—for which the image quality has improved markedly since the HIP Breast Project. This improvement almost certainly accounts for the dramatic rise in the incidence of ductal carcinoma in situ in the last decade.[12] Although changes that enhance detection are generally interpreted as being improvements, their impact on patient outcomes are unpredictable. Lower detection thresholds may increase detection of clinically relevant cancer and enhance the benefits of screening, but they may also increase detection of pseudodisease (subclinical disease that would not become overt before the patient dies of other causes) and dilute the benefits of screening through unnecessary therapy.

ARE THE RESULTS OF THE STUDY VALID?

After deciding that a study is applicable to one's own practice, the clinician will want to assess the validity of its results. In short, the question is: Are the findings an unbiased estimate of the effectiveness of screening? In this article we assume that the study is randomized, perhaps the most important prerequisite for producing an unbiased estimate.

Were the Groups Similar at the Start of the Trial?

Participants are generally randomized to either receive an invitation to be screened or not (or alternatively randomized to receive screening of varying frequency). Readers can be reassured about the adequacy of randomization by considering the similarity of the two (or more) groups, generally reported by investigators in a table of baseline data regarding study participants. These usually include age, gender, and other known or suspected risk factors for disease. Randomization helps to ensure that known and unknown risk factors are evenly balanced.

In considering the similarity of the groups, one should not compare data collected after the study commences—in particular, the number of cases of the target disease. For example, an increased number of lung cancer cases in the screened group of the Mayo Lung Study has been recently been interpreted as a failure of randomization.[13,14] This interpretation is flawed as increased detection is expected. In fact, if screening is to work, more cases must be detected.

Were Outcome Assessments "Blinded" to the Randomization?

Unlike randomized trials of therapy, "blinding" is not complete in screening trials as both patients and provid-

Table 2. Questions About the Results of Three Randomized Trials of Screening Strategies Intended to Reduce Cancers of the Lung, Breast, and Colon

	Mayo Lung Study		HIP Breast Project		Minnesota Colon Cancer Control Project*	
Target disease	Lung cancer		Breast cancer		Colon cancer	
Screening strategy	Chest x-rays three times a year		Annual mammography (two-view)		Annual fecal occult blood testing	
Study participants	Male smokers over age 45 ($N \approx$ 10,000)		Females age 40–64 ($N \approx$ 60,000)		Males & females age 50–80 ($N \approx$ 30,000)	
	Screened	**Control**	**Screened**	**Control**	**Screened**	**Control**
Results about the test						
Are more cases of disease detected in the screened group early in the trial?	After 6 years: 206	160	After 4 years: 250	219	After 5 years: ~115	~115
Are fewer cases of advanced disease detected over the course of the trial?	Unresectable cases: 112	109	Node + cases: 97	125	Stage D cases: 36	68
Results about the treatment following the test						
Has advanced disease been prevented in the screened group?†	Lung CA mortality: 32	30	Breast CA mortality: 3.2	4.4	Colon CA mortality: 4.5	6.6
Is it clear that early detection and treatment produces a net gain?†	All-cause mortality: 235	231	All-cause mortality: 73.7	75.4	All-cause mortality: 183	183

**For simplicity, the biennial screening strategy is excluded from this table. Only the results from annual screening and control patients are reported.*

†Disease-specific and all-cause mortality rates are per 10,000 person-years of observation. The length of follow-up is approximately 8, 10, and 13 years for the three trials, respectively. Mortality rates for the HIP Breast Project and the Minnesota Colon Cancer Control Project were calculated based on the reported number of deaths and person-years in the study.

ers need to know the results of the screening tests (and hence the randomization). Because providers are not blinded, subsequent treatment should be standardized for different stages of disease. Blinding should be expected of certain study personnel, however, particularly those making judgments about intermediate and long-term outcomes (e.g., diagnosis, staging, cause of death).

Death due to the target disease (i.e., disease-specific mortality) is generally the primary outcome in screening trials. Because these deaths are often rare, the appropriate attribution of the cause of every death is critical. Among the approximately 4,500 deaths in the HIP Breast Project, for example, the difference between the screened and control groups was merely 38 breast cancer deaths,[10] making the question of whether a patient died of breast cancer or other causes extremely important. Consequently, the investigators developed explicit rules detailing how deaths should be classified. To remove the potential for bias, furthermore, investigators assessing the cause of death were also blinded to the study group assignment.

Were the Subjects Analyzed According to Randomization?

Compliance in randomized controlled trials of screening is rarely complete. Some patients invited to screening may not accept while some not invited may undergo screening anyway. Lack of compliance in either group decreases the differences in the outcomes of the two groups. Nevertheless, because some lack of compliance is expected in a population invited to screening, it is appropriate to compare the outcomes of two groups directly if the purpose of the study is to estimate the effect of an invitation to screening. However, if the purpose of the study is to estimate the effect of accepting the screening invitation, then some adjustment must be made for the lack of compliance.[15] In the HIP study, for example, compliance among the screening group was only 67%. If all those invited to screening had accepted, the mortality reduction from screening would have been about 43% instead of the reported 29%.

WHAT WERE THE RESULTS?

Randomized trials of screening are actually a hybrid investigation, the results of which are a function of both the diagnostic accuracy of the test and the effectiveness of advancing the time of treatment. The ultimate study question is whether the patients in the screening limb somehow do better. The answer to this question is a function of both the test and the treatment.

This test-treatment intervention provides two opportunities in which the results can be evaluated. Preliminary reports from a randomized trial will focus on the question about the test: "Does early detection actually occur?" This is a necessary, although insufficient, requirement for screening to be effective. Subsequent reports will focus on the question about the treatment: "Does therapy following early detection provide benefit to the population being screened?" Ultimately there needs to be a full accounting of the outcomes of both those correctly diagnosed by screening and those incorrectly diagnosed (with the side effects experienced by subjects with false-positive results being most relevant).

Critical readers should consider both the test and treatment elements when reviewing a randomized trial of screening. Some questions are relevant to the test; others, to the treatment following the test.

Results About the Test

Are More Cases of Early Disease Detected in the Screened Group?

The ability of the screening test to detect disease early ought to be evident in the first round of screening in the trial.[16] This increased rate of detection is necessary for screening to work. The proposed mechanism for benefit is familiar to us all: find disease early and treat it promptly. Finding disease earlier in the screened group than in the control group means more will have been found in the screened group at any point in time.

To illustrate this phenomenon, Figure 1 displays cumulative incidence over time (idealized for simplicity). Those cases only detectable by screening are defined as subclinical; those cases evident by signs and symptoms are defined as overt (see Appendix A for a glossary of terms). The left-hand portion shows an initial "jump" in cumulative incidence, as prevalent subclinical cases (existing prior to randomization) are rapidly detected in the screening limb. In the control limb, cases must progress and become overt to be detected. Because the underlying rate of disease initiation is equivalent, the rate of disease detection soon equalizes in the two groups (i.e., the lines are parallel). But as long as screening continues, the line for the screened group is shifted to the left. Thus, at any point in time, the screening limb will have accumulated more cases from the time of randomization. This shift reflects the earlier detection in the screened group and is known as the lead time.

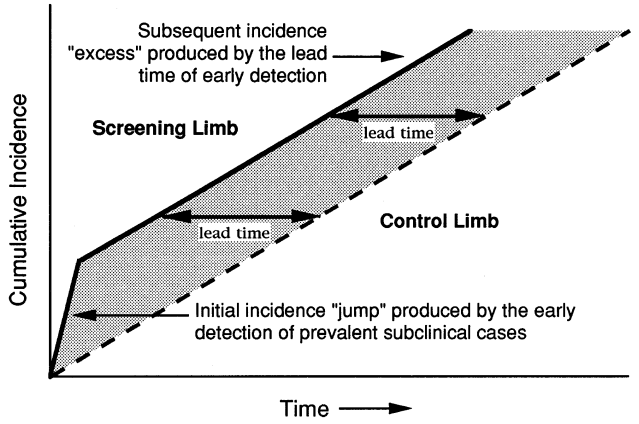

FIGURE 1. Increased detection of disease in the screening limb early in the trial. This example assumes that prevalent subclinical cases remain in both limbs and that screening occurs repeatedly. The horizontal distance between the two curves is the lead time.

It is important to emphasize here that lead time is potentially a good thing. Finding disease early is the primary rationale for screening. Thus, the suggestion that a lead time exists, is a suggestion that the test is working. The negative connotation associated with the term (lead time bias) refers to failing to account for lead time when measuring survival (as discussed below).

Table 2 shows that more disease was found in the screening group in both the Mayo Lung Study and the HIP Breast Project. The equal number of cases detected in the Minnesota Colon Cancer Project would seem to suggest that screening could not work—that screening did not detect more cases of disease. But this more likely reflects the classification of disease in the colon: where polyps may be considered "precancerous" and not cancer. The combination of frequent polyp detection in the screened group (in the thousands) and the subsequent evidence demonstrating long-term benefit assures the reader that more precancerous disease was detected in the screened group.

Recent reports have questioned the validity of the Mayo Lung Study because the imbalance in cumulative lung cancer incidence persisted 3 years after the cessation of screening.[13,14] Rather than accepting an implication of faulty randomization (leading to an "imbalance of coexisting risk factors"), the critical reader should understand that there are a number of reasons why increased incidence in the screening limb can persist despite the cessation of screening. First, the follow-up may be too short and the imbalance explained by subclinical cases not yet identified (i.e., not yet overt) in the control limb. Second, there may be pseudodisease in the screening

limb—cases of disease that would have never become evident in the control limb. Finally, participants previously exposed to intensive screening may continue to aggressively pursue screening. Although waiting out the lead time will correct the imbalance in the first case, no amount of follow-up will correct the imbalance in the latter two.

Are Fewer Cases of Advanced Disease Detected in the Screened Group?

Although increased detection early in the trial is the first available evidence that early detection is occurring, a reduction in advanced disease in the screening limb over the course of the trial is a more important finding.[16] Both the HIP Breast Project and the Minnesota Colon Cancer Control Project had very favorable results with regard to the number of cases of advanced disease. The number of patients with breast cancer whose axillary nodes tested positive at diagnosis was 20% lower among those screened. The number of patients with colon cancer whose disease was classified as Duke's stage D at diagnosis was almost 50% lower. The number of patients with unresectable lung cancer in the Mayo Lung Study, however, was virtually equivalent in the two limbs—an important indication of the failure of the screening test.

The finding of fewer cases of advanced disease at the time of diagnosis in the screening limb helps confirm that early detection is occurring. Once again, however, this finding is a necessary, but not sufficient prerequisite for a useful screening program. To complete the case for screening requires that early treatment be effective.

Results About the Treatment Following the Test

Has Advanced Disease Been Prevented in the Screened Group?

The most convincing evidence that disease progression is slowed by early detection and treatment is the demonstration of a reduced burden of advanced disease at the end of the trial. This is most indisputably measured in terms of deaths from disease. Although this idea is conceptually simple, it is easy to be confused about what is the right mortality measure.

Case fatality (and its complement, case survival) are measured from the time of diagnosis. These measures are appropriate for treatment trials because diagnosis occurs before randomization. However, these measures are inappropriate for screening trials because diagnosis occurs after randomization. In fact, to the extent that screening advances the time of diagnosis, these measures are biased in favor of screening. Three distinct biases affect the often quoted comparison of case fatality in screen–detected versus clinically detected cases of disease: lead time bias, failure to adjust for the earlier diagnosis in the screened group; length bias, failure to adjust for the disproportionate selection of slowly progressive disease in the screened group; and overdiagnosis bias, failure to adjust for the detection of pseudodisease in the screened group (subclinical disease that would not become overt before the patient dies of other causes).[4] As shown in Table 3, their combined effect can be powerful as significant benefits of screening were suggested by this measure in three widely acknowledged negative screening trials.[9,17,18]

Table 3. A Comparison of Case Fatality and Disease-Specific Mortality

	Case Fatality	Disease-Specific Mortality
Measure description		
Question addressed	What are the outcomes for those with disease?	What are the outcomes for those who are screened?
Numerator	Number of deaths from disease	Number of deaths from disease
Denominator*	Number of individuals diagnosed with the disease	Number of individuals in the study group as a whole
Starting point for measurement	Time of diagnosis	Time of randomization
Appropriate use	Trials of therapy (where patients are randomized *after* diagnosis)	Trials of screening (where patients are randomized *prior to* diagnosis)
Three examples of bias when case-fatality is applied to screening†	(Screened vs Control)	(Screened vs Control)
Mayo Lung Study[9]	60% vs 85%	No difference
Czech Lung Cancer Screening Trial[17]	77% vs 100%	No difference
Malmö Mammographic Screening Trial[18]	3% vs 15%	No difference

**When calculating a rate (i.e., case fatality rate, disease-specific mortality rate), the denominator is expressed in person-years (i.e., the sum of the number of years each individual is at risk).*

†These are three examples of the spurious effects of lead time and length biases when case fatality is used.

Disease-specific mortality avoids these problems by using an unbiased denominator (the study group as a whole) and an unbiased starting point (time of randomization). It is calculated as the number of patient deaths from the disease divided by the "person-years" in which the study group is at risk. As shown in Table 2, disease-specific mortality was slightly higher in the screened group in the Mayo Lung Study. In the case of the HIP and Minnesota studies, patients who were screened did experience significantly lower disease-specific mortality.

Although reductions in disease-specific mortality are conventionally expressed in relative terms, many have argued that a measure of absolute risk reduction is important to place the reported benefit in perspective.[19] Reports that screening reduces breast and colorectal cancer deaths by 30%, for example, distract clinicians from the absolute reduction demonstrated in Table 2—about 1 breast cancer death and 2 colorectal cancer deaths per 1,000 persons screened over 10 years. Nevertheless, the finding of lower disease-specific mortality in the screened group can be viewed as a valid measure of benefit using the relevant and relatively unambiguous outcome of deaths from disease.

Is It Clear That Early Detection and Treatment Produces a Net Gain?

Finally, it is worth considering whether death from disease is a too narrowly focused outcome measure. Screening may also have unintended adverse effects (stemming from the test itself, subsequent diagnostic efforts, or therapy) that are not included in measures of disease-specific mortality.[16] The measurement of cardiovascular mortality, for example, following the administration of lipid-lowering agents in primary prevention would miss any effect on other causes of mortality.[20] Even rare adverse effects may be significant as they can accrue not only to those who have disease, but also to the larger group who test positive. To consider whether a reduction in disease-specific mortality is somehow offset by an increase in mortality from other causes, all-cause mortality must be examined.

There is no evidence of any such tradeoff in the HIP Breast Project. As shown in Table 2, all-cause mortality was lower for those who were screened. The reduction in all-cause mortality, in fact, nicely mirrors (i.e., has a similar magnitude to) the reduction in disease-specific mortality. Thus, one can be reassured that screening did not have an unexpected adverse impact on mortality. In the Minnesota Colon Cancer Control Project, however, the reduction in deaths from colorectal cancer was accompanied by a comparable increase in deaths from ischemic heart disease, resulting in identical all-cause mortality.[11] It is reasonable, therefore, to wonder whether the frequent use of colonoscopy following fecal occult-blood screening somehow increases cardiac mortality sufficient to offset the reduction in mortality from colorectal cancer.

Although all-cause mortality may be quite insensitive to the beneficial effect of screening (particularly when the disease is rare), there are two arguments for examining it along with disease-specific mortality. First, all-cause mortality helps ensure that a major harm (or benefit) is not being missed. It is all inclusive and provides data relevant to the question of whether other risks are somehow changed by the test-treatment strategy. Second, all-cause mortality provides an important perspective on the magnitude of benefit. It puts disease-specific mortality reduction in the context of other competing risks. This helps the prospective screenee focus on the overall benefit that can reasonably be expected. Although one should not expect statistically significant changes in all-cause mortality (given sample size constraints), its role in generating hypotheses about unexpected risks and in providing perspective should not be ignored.

Evaluating the applicability and validity of randomized trials of screening has many similarities with evaluating randomized trials in general. Evaluating the results, however, raises some issues peculiar to screening. First, readers should recognize that finding more cases of disease in the screening limb is to be expected if screening is to work. Furthermore, regardless of the real effect of earlier diagnosis, they should understand that it improves the stage distribution at the time of diagnosis and case fatality (and hence survival) from the time of diagnosis. Finally, readers should remember that, even if screening is effective, there are side effects for those who do not have disease: false-positive test results producing anxiety and overdiagnosis leading to unnecessary treatment.

We are indebted to Robert Greenberg, Lisa Schwartz and Steve Woloshin, colleagues at Dartmouth who provided valuable critique. In addition, we deeply appreciate the reviews given us by Noel Weiss at the University of Washington.

REFERENCES

1. US Preventive Services Task Force. Guide to Clinical Preventive Services. 2nd ed. Baltimore, Md: Williams and Wilkins; 1996.
2. Wilfond BS, Nolan K. National policy development for the clinical application of genetic diagnostic technologies: lessons from cystic fibrosis. JAMA. 1993;270:2948–54.
3. Russell LB. Educated guesses: making policy about medical screening tests. Berkeley, Calif: The Regents of the University of California; 1994:128.
4. Black WC, Welch HG. Advances in diagnostic imaging and overestimations of disease prevalence and the benefits of therapy. N Engl J Med. 1993;328:1237–43.
5. Cochrane AL, Holland WW. Validation of screening procedures. Br Med Bull. 1971;27:1–8.
6. Lee JM. Screening and informed consent. N Engl J Med. 1993; 328:438–40.
7. Guyett GH, Sackett DL, Cook DJ. User's guide to the medical literature. How to use an article about therapy or prevention. JAMA. 1993;270:2598–2601.
8. Fontana R. Screening for lung cancer. In: Miller A, ed. Screening for Cancer. Orlando, Fla: Academic Press; 1985;377–95.

9. Sanderson D. Lung cancer screening: The Mayo Study. Chest. 1986;(4 suppl):324S.
10. Shapiro S, Venet W, Strax P, Venet L. Periodic Screening for Breast Cancer: The Health Insurance Plan Project and Its Sequelae 1963–1986. Baltimore, Md: John Hopkins University Press; 1988.
11. Mandel JS, Bond JH, Church TR, et al. Reducing mortality from colorectal cancer by screening for fecal occult blood: Minnesota Colon Cancer Control Study. N Engl J Med. 1993;328:1365–71.
12. Ernster VL, Barclay J, Kerlikowske K, Grady D, Henderson IC. Incidence of and treatment for ductal carcinoma in situ of the breast. JAMA. 1996;275:913–8.
13. Strauss G, Gleason R, Sugarbaker D. Screening for lung cancer re-examined. A reinterpretation of the Mayo Lung Project randomized trial on lung cancer screening. Chest. 1993;(4 suppl):337S–41S.
14. Strauss G, Gleason R, Sugarbaker D. Chest X-ray screening improves outcome in lung cancer. A reappraisal of randomized trials on lung cancer screening. Chest. 1995;(6 suppl):270S–9S.
15. Black WC, Welch HG. Screening for disease. Am J Roentgenol. 1997;168:3–11.
16. Morrison AS. Screening in Chronic Disease. 2nd ed. New York, NY: Oxford University Press; 1992:254.
17. Kubick A, Parkin D, Khlat M, Erban J, Polak J, Adamec M. Lack of benefit from semi-annual screening for cancer of the lung: follow-up report of a randomized controlled trial on a population of high risk males in Czechoslovakia. Int J Cancer. 1990;45:26–33.
18. Andersson I, Aspegren K, Janzon L, et al. Mammographic screening and mortality from breast cancer: the Malmo mammographic screening trial. BMJ. 1988;297:943–8.
19. Naylor C, Chen E, Strauss B. Does the method of reporting trial results alter perceptions of therapeutic effectiveness? Ann Intern Med. 1992;117:916–21.
20. Muldoon MF, Manuck SB, Matthews KA. Lowering cholesterol and mortality: a quantitative review of primary prevention trials. BMJ. 1990;301:309–14.

APPENDIX A
Glossary of Terms Relevant to Screening

Screening: the systematic examination of those who are apparently well (or who are apparently free of the target disease) to identify and treat subclinical disease (or even predictors of future disease).

Overt disease: a condition in which disease is detectable by signs and symptoms.

Subclinical disease: a condition in which disease is detectable by testing but is not evident by signs or symptoms.

Pseudodisease: subclinical disease that would not become overt before the patient dies of other causes.

Disease spectrum: the full range of disease extent; from precursor states (e.g., early disease) to a fulminating, florid condition (e.g., advanced disease). This term can also encompass the various rates of disease progression.

Lead time bias: overestimation of survival duration among screen-detected cases (relative to those detected by signs and symptoms) when survival is measured from diagnosis. This is simply a reflection of earlier diagnosis.

Length bias: overestimation of survival duration among screen-detected cases due to the relative excess of slowly progressing cases. These are disproportionally identified by screening because the probability of detection is directly proportional to the length of time during which they are detectable (thus inversely proportional to the rate of progression).

Overdiagnosis bias: overestimation of survival duration among screen detected cases due to the inclusion of cases of pseudodisease.

Case-fatality rate: the rate of death among cases with a specified diagnosis, measured from the time of diagnosis.

$$\text{Case-fatality rate} = \frac{\textit{Number of deaths from a disease (in a given period)}}{\textit{"Person-years" in patients diagnosed with the disease (in the same period)}} \times 100$$

Disease-specific mortality rate: the population based rate of death from a specified disease. In a randomized clinical trial, this is measured from the time of randomization.

$$\text{Disease-specific mortality rate} = \frac{\textit{Number of deaths from a disease (in a given period)}}{\textit{"Person-years" in population at risk (in same period)}}$$

All-cause mortality rate: the population-based rate of death. In a randomized clinical trial, this is measured from the time of randomization.

$$\text{All-cause mortality rate} = \frac{\textit{Number of deaths (in a given period)}}{\textit{"Person-years" in population at risk (in same period)}}$$